DOCUMENT RESUME

ED 192 621 FL 011 888

AUTHOR Cargill-Power, C.
TITLE Cultural Bias in Testing ESL.
PUB DATE Feb 80
NOTE 12p.; Paper presented at the Annual Meeting of the Southeastern Conference on English in the Two-Year College (15th, Birmingham, AL, February 21-23, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; Cultural Context; Culture Conflict; *Culture Fair Tests; *English (Second Language); *Language Tests; *Test Bias; Test Items

ABSTRACT

Although cultural content is unavoidable as a backdrop for good language testing, cultural bias in testing English as a second language presents many dangers. A picture cue calling for a correct grammatical response may evoke an incorrect answer if the pictorial content is culturally coded. The cultural background behind a test must be accurately reflected, without misinformation or misleading emphases. Current events, especially those involving geography and "hot news items," should be questioned as a cultural setting for test items. A test designed for use abroad should not contain culture-coded items reflective only of certain American households. Test administration must take into account cultural conditioning that may find the student terrorized in the face of a typographical error, a timed test, or the language-specific bias of a cloze test. Syntactic differences between the native language and English may render certain tests useless as placement instruments. Test formats that are taken for granted in the American educational system, especially cloze tests, should be examined for the degree to which they may, by their nature, handicap one linguistic group or another. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Cargill-Power
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED192621

FL011888

"Cultural Bias in Testing ESL"

C. Cargill-Power

There are potential dangers in culture coding tests. While language instruction should not be devoid of cultural content, cultural bias and culture coding may yield unreliable test results. There are various ways in which a test may be culturally biased. The culture-coded test is a test which requires the student to know something outside the intended forms of the test. That is, its validity is questionable because some extra-linguistic information is required of the student in order to answer.

An example of the danger from such material could be a placement test in which oral production is tested by means of a picture stimulus. The picture cue calling for the correct grammatical response is preceded by a different picture cue with a similar scene. For example, the student taking this test first sees a picture of a boy smiling. The student is told, "the boy is happy.". The student is then shown another picture cue with a girl smiling. If he is expected to answer, "The girl is happy", the test is heavily culture coded; it would certainly be of doubtful reliability in assessing the student's ability to translate that picture cue into the idea that the person in the picture is happy. We know in American culture that a smiling face usually means that a happy person is wearing it. However, experience in the Vietnamese integration programs going on around the country shows

that a smiling Vietnamese boy may not be happy; he may be embarrassed, confused, or even angry. (The famous Vietnamese smile!)

Fortunately, this particular test was developed by experienced teachers whose aim was not to elicit only the response, "The boy is happy" which would have been culturally unfair. In fact, this test was structurally controlled. The sought-after response was merely a form of the verb "to be" in the present tense. The acceptable responses have a wide range: (Marie DeShields Jawardi 1976)

1. The boy is happy.
2. The boy is smiling.
3. He is happy.
4. He is smiling.
5. The boy's happy.
6. The boy's smiling.
7. He's happy.
8. He's smiling.

Where the test makers avoided culture coding was in allowing even such answers as: "He's tired." Although the student might not interpret the smile as a sign of happiness, he could demonstrate that he knew and could handle the present tense, third person singular form of the verb "to be".

What are the pitfalls relating to culture which one should avoid in language teaching and testing? In her presentation at the International Seminar on Language Testing (1973), E. Condon stated that culture teaching in the classroom should include facts on civilization such as history, geography, and the arts. Culture, however, can be found not only in such documented evidence as historical reading selections but also in what

Condon calls explicit and implicit aspects of civilization. Explicit forms of culture are represented by class distinctions, family relationships, housing practices, and eating habits. Implicit cultural information comes from culturally conditioned factors such as unspoken attitudes and beliefs, she says.

This information may be without value in test making if the item may be true in some regions or local communities but may not be widespread enough to be justified in a test. Of course, references to local customs are encouraged for use in local teacher-made tests. However, cultural bias on the part of the teacher in the testing situation should be avoided. Today, with the highly mobile society we all live in, we are encouraged to situationalize teaching and testing experiences according to the realities to which the students will have to relate. That is, our cultural input must be controlled for its relavance.

It must also be controlled for its accuracy in general. For example, the Comprehensive English Language Test (1970) included out of 35 items only 11 which were neutral. The others showed American society characterized by rejection, anger, fear, and violence. While all these appear in American life, the inclusion of so many cannot but influence the student's picture of the U. S. in a more negative way than is desired and stimulate reflection on these topics or reaction against them.

In our text books, from which many teachers draw their test material, at least in order to control valid vocabulary for test items, there may be out-and-out errors in reference to cultural background. While these items may have been correct at the time of the writing of the text, they no longer are accurate. For example, a statement that the Empire State

Building is the tallest building in the world would have to be changed if accuracy in cultural content is desired. On the other hand, if the structural form in focus is the superlative form "tallest", then we might say that the misinformation given serves as a distractor, but we will nevertheless receive the correct grammatical form.

Sometimes, culture coding serves as more than a distractor; it serves as the main attraction. For example, in a well-known text in use today, there is an example which requires the student to know cultural background in order to answer with the correct structural feature. The grammatical point in focus is the comparative form. The sentence reads:

> "Davis has an interest in medieval art. He has been to both the Cloisters and the Museum of Modern Art. Naturally, the Cloisters are _____.
> (more interesting)

This item requires the student to know that:

1. The Cloisters is/are a museum.
2. It contains medieval art.

A similar item requires the student to know Yankee Stadium, the Polo Grounds, and Ebbetts Field.

Ideally, of course, all informed and experienced test writers would have done an item analysis after the test and would have concluded that the index of difficulty is too high. The item would have been stricken.

Accuracy may be lost also not only from dated textbooks but also from dated information which the teacher-testmaker possesses. For example, to refer once again to the issue of the tallest building, how many ESL teachers could accurately answer the question which required the superla-

tive form in the item:

The Sears Tower in Chicago is tall.

The World Trade Centers are tall.

In this example, there is no grammatical cue; therefore, the answer can only come from a cultural cue which the student has gotten outside the class (or at least outside the test.)

Another possible area for teacher inaccuracies is in the realm of geography. The world's borders are changing in some areas so fast that the teacher would be well advised to double check his information with the morning newspaper. His cultural input in his classroom exercises and tests may be biased by his own ignorance.

Appropriateness of "hot news items" as a cultural setting for a test item should be questioned. For example, some text writers have shown a lack of sensitivity to such an event as the assassination of President Kennedy. Is this information really appropriate as a backdrop for a test item? Surely the painful event in Ford's Theater a century ago was probably avoided in other classes for a much longer period of mourning than would be expected in history classes. The use of such hot items as a backdrop for a grammatical exercise or test is open to criticism. It would not be condemned, of course, in a reading test for cultural background.

Another way in which culture coded information on a test may produce unreliable performance on the part of a test taker is where a test item describes a setting which may be typical of some American homes but which would be unheard of in the home country of the international student.

For example, a test administered for Spanish speaking students in which a picture cue is introduced showing a boy volunteering to help his mother with the housework is not acceptable. One might argue that, after all the student studying here in the U. S. should get used to such ideas since they are not totally unheard of here. However, the picture changes when this test item appears in a commercially produced test and is shipped to South America. The student, upon seeing this scene, is immediately hit by culture shock. Is this healthy in the testing situation? What toll does it take from his test score?

Another example of cultural interference is where the test administration itself takes a toll in the score. For example, in most of our E.S.L. classrooms, a student is allowed and even urged to ask questions during the test if he is too puzzled to go on. This is usually done without disturbing other students through a simple raise of the hand. As most of us know, frequently a student will discover that last of several typographical errors which we, the testmakers, have tried desperately to find before test time. In this case, the American student would simply make the inquiry, and the correction would be noted for the benefit of the whole class. (Fortunately, on standardized tests, such kinks would have already been removed.) The freedom to ask questions during a test is absent in many cultures. A foreign student would rather leave a section blank, thus taking a chance on failing the test, before he would venture to ask a simple question.

Culture shock may result from another aspect of test administration: the timing of the test itself. Most international students will not have experienced the timed test until the Test of English as a Foreign Language

(TOEFL). The reliability factor in the student's performance is again somewhat in question when he sits nervously ever aware of the clock.

Culture coding also appears from the administrative point of view in the item types themselves. This is a problem that appears in the item types known as multiple choice and cloze; as with the timed test, the student is at least temporarily confused, even given a battery of instructions explaining what a multiple choice test requires of the student to indicate his choice.

Cloze testing has been used as an instrument to determine placement into a language program, proficiency in a language, reading readiness, and readability of material for international students.

Alderson defines the term cloze procedure as being used in three different ways. The first and most general definition is "the systematic deletion of words from text", where "systematic" remains undefined. The second definition takes the word "systematic" and devides it into two types of systems: that is, either a random (or, better, pseudo-random) deletion of words, or a rational deletion. A third definition, which is increasingly common in the literature, is the deletion of every fifth word from the text (that is, not just pseudo-random, but a specific deletion frequency). (Alderson 1979).

The scoring of a cloze test may vary. The most common way is to accept as correct only an answer which gives the exact word which had been deleted from the text, ignoring misspellings. This is the most rigid and, in my opinion, the one which tests memory span rather than comprehension. In most ESL testing situations using cloze, the exact word is <u>not</u> sought, but rather an appropriate synonym may be acceptable to show

that the reader has comprehended the message. This seems to be the least biased procedure which is, at the same time, neither too rigid nor too flexible. An example of the former might be where any item of the same form class is acceptable even though it is not logically related to the passage (form class here means the same grammatical category or part of speech). This seems to me objectionable because it may reliably test the reader's perception of the morphological or syntactic rules of the language but not even begin to measure his comprehension of the text. For example, there are many definitions of a noun, such as the traditional one, that it is the name of a person, place, or thing. Modern linguistics has little argument with this, except that it merely defines the single word noun, not the noun phrase or noun clause which as a whole may act like a noun. We add syntactic definitions to this tradionally lexical one, such as: a noun is a word which can be directly modified by the definite article, "the". Such a definition is useful in computer translations and generative grammars where not only lexical or semantic features are given but syntactic features also are necessary. An example of this would be the word "Irish", which traditionally is considered an adjective. However, when we place the definite article in front of "Irish", we get "the Irish", which is a noun phrase with an adjective funtioning as a noun.

The cloze test which is so loose that it admits as correct words of the same form class does not discriminate between those who have comprehended what they have read and those who recognize the syntactic cue to the correct form class. The former selects a word which fits in to the logic or the stream of thought in the passage, while the latter gets the same credit for having only a small amount of information about the structures he reads, and he answers with an inappropriate noun in a space following the word "the".

Cultural bias comes in when the test is being used to assess the proficiency of a non-native speaker or language X. It also comes into play when the test is used for placement purposes. For example, the student may exhibit a response which does not fit the format or the cloze procedure. The test solicits single word responses in the deletion slot. Bias in test format may appear in testing such groups as Thai students. Their language does not mark work boundaries as we know them. Therefore, Thai students tend to make more errors on cloze tests than do students from other language groups because Thai's tend to fill in the blanks with multiple word entries (Oller, 1972). In doing so, they cannot even get credit for the correct form class since the syntactic structures they give are so complex that they obscure any identification of form class.

We know from comparisons between these cloze tests and other standardized tests designed for proficiency or placement that there is a low correlation between them and cloze when used with these students. Which one is unreliable? The cloze is, obviously.

What can be done to avoid these problems? As we know from experience with American students taking the SAT's they are also victims of nervousness the first time around. We routinely advise them to take the exam in their junior year in order to give them some familiarity with a timed, multiple choice test. Should we not advise our international students to do the same? Of course, many of our students do take the test more than once. This is a highly advisable approach. The test should test what it is intended to test; otherwise, it loses validity and reliability . Tests which show the student's ability to perform and manipulate the <u>format</u> first and the <u>content</u> second should be avoided.

In summary, cultural content is necessary in language instruction. it is unavoidable as a backdrop for good language testing. It should at least be free of misinformation, teacher bias, and emotionally traumatizing topics. Culture coded input both in standardized and classroom constructed exams should be taught before tested if that is to be part of the content required. Culturally conditioned factors such as those in test administration should be carefully considered when an overall judgement of student performance is desired. Perhaps, for the test taker who has never experienced a timed, multiple-choice test, it should be strongly recommended that he plan to take the test more than once. The examiner should be aware of culture-coded test formats such as cloze in order to avoid those which by their very nature may handicap one linguistic group or another.

My own research in the English Language Center is seeking to add to our understanding of the reliability of cloze testing. I am attempting to identify patterns of intercultural interference which may prevent specific language groups from doing well on such tests. I am looking at responses from students from many language groups such as those from the Orient, the Middle East, and Latin America.

Two types of cloze tests are under study. One is a strictly random deletion of every 7th word; the other involves deletion of only one form class, that of preposition. The scores received by the student body are then correlated with the CELT, a widely recognized standardized examination.

The results of these tests are examined from the point of view of the level of proficiency and cultural ethno-linguistic background.

We are also preparing a vocabulary placement examination to supplement the CELT vocabulary test which correlates very poorly with the listening and structure sections. This vocabulary test in context is to be for advanced level V students in the Center to determine their readiness to go on to academic programs.